Comparison of the Accuracy of Several Systems For Hospital Quality Screening and Assessment

Final Report to the
Health Care Financing Administration
Office of Research and Demonstrations

Grant #17-C90099/3-02

Submitted by

Haya R. Rubin, M.D., Ph.D.
Principal Investigator
Assistant Professor of Medicine and Health Policy and Management
Director, Quality of Care Research
Johns Hopkins Program for Medical Technology and Practice Assessment
1830 E. Monument St. #8015
Baltimore MD 21205

May 1, 1995

Comparison of the Accuracy of Several Systems For Hospital Quality Screening and Assessment

Grant #17-C90099/3-02

Final Report

Table of Contents

EXECUTIVE SUMMARY 2-4

BACKGROUND 4

PURPOSE AND AIMS 5

METHODS 5-9

RESULTS 9-13

CONCLUSIONS AND RECOMMENDATIONS 13-15

REFERENCES 16

TABLES (1-36)

APPENDICES (A1-A3) 1-25

Abbreviations used in Appendix Tables 26

EXECUTIVE SUMMARY

Background: The Health Care Financing Administration contracts with peer review organizations (PROs) to obtain physician review of medical records. While the results of individual case review by one or two physicians have been recognized to be unreliable, and HCFA no longer contracts with PROs to act on results of individual reviews, results of review of groups of patients treated by a given physician or provider are used by HCFA to help physicians and providers identify quality improvement opportunities.

To make the review process more efficient, it has seemed desirable to use a screening strategy to decide which records should be referred to a physician, thus utilizing more expensive physician time most efficiently. Such screening systems have received little evaluation and rarely have alternatives been directly compared. HCFA is developing the Uniform Clinical Data Set; one function of this data set was to provide automated algorithms to screen cases for whether they should be referred for review.

Aims: The purpose of this study was to measure the sensitivity, specificity and predictive value of three different medical -record based screening systems for quality of care problems experienced by patients with coronary artery disease hospitalized for myocardial infarction (MI), percutaneous transluminal coronary angioplasty (PTCA) or coronary artery bypass graft (CABG) surgery: The three screening systems were:

1) an early prototype of the Uniform Clinical Data Set (UCDS) that was being pilot tested by several peer review organizations at the time the study was undertaken;

2) Johns Hopkins hospital-wide (JHH) screens, including HCFA generic screens that were used by many peer review oganizations

3) the Harvard Medical Practice Study (HMPS) screening criteria,

The study originally proposed to measure the screening systems against two different reference standards , blinded to the screening results: 1) agreement by at least two reviewing cardiologists that there was a quality of care problem, 2) the determination that an adverse event was due to negligence, as defined in the Harvard Medical Practice Study.

Methods:

A stratified sample of 460 medical records was drawn in three strata: myocardial infarction, coronary angioplasty, and coronary artery bypass surgery, from admissions in the period 7/1/89-6/30/91 at Johns Hopkins Hospital. Within each cohort, the sample also was stratified by whether or not records had passed Johns Hopkins Hospital's hospital-wide quality screens. Records were screened using all three screening systems by trained quality assurance nurses. All records were also reviewed by two to three cardiologists using structured implicit review instruments to determine the reference standards. All records were also coded for APACHE-III severity of illness scores and the Charlson comorbidity index. ROC-like curves were obtained using different subsets of screens and the area under the curve compared for these different subsets. Nine records were unavailable for review and twenty-eight were reviewed by only one physician and were excluded from the study results that use two physician reviews.

Results:
Agreement among physicians about whether there was a quality problem or not was 76% (kappa 0.37). Both physicians agreed that there was a quality problem in 17.8% of cases, pooling all three cohorts, and adjusting for the stratification procedures. In all three screening systems a small number of screens detected most of the quality problems. Overall agreement in the nurse screening result about whether to refer a record for physician review was 98% for UCDS, 77% for JHH ,and 85% for HMPS . The systems performed as follows in detecting quality problems agreed upon by at least two of the cardiology reviewers:

	Sensitivity	Specificity	Positive Predictive Value	MD hrs per detected problem	RN hrs per detected problem
UCDS	85%(+/-16)	48%(+/-12)	25%(+/-11)	2.1	6.7
JHH	74 %(+/-12)	57% (+/-11)	27 (+/-12)	1.9	1.9
HMPS	69%(+/-16.3)	56%(+/-12)	26(+/-11)	2.0	2.0

Sensitivity was similar for all cohorts for UCDS, and better than JHH and HMPS for all. Sensitivity for MI patients cannot be evaluated properly due to the absence of quality problems making it impossible to produce pooled estimates in that cohort. Specificity of JHH and HMPS screens exceeded that of UCDS for PTCA and MI patients. Positive predictive values for UCDS were similar for all cohorts, and similar to that of JHH and HMPS screens, except for the MI cohort, which again due to the absence of quality problems in one stratum made pooled estimates difficult to interpret. Because of this, the number of physician and nurse hours per quality problem detected was greater with JHH and HMPS screens than for UCDS for MI patients (physician hours: JHH 6.3, HMPS 7.1, UCDS 1.9, nurse hours: JHH 9.7, HMPS 9.6, UCDS 5.6). For all other cohorts, nurse hours per quality problem detected were 6.7 hours for UCDS and 1-2 hours for the other systems, while physician time per quality problem was approximately 2 hours for all systems.

Conclusion:
For the cardiac patients at this hospital, the version of UCDS tested was more sensitive and less specific for detecting quality problems than the other two screening methods, but was much less efficient per quality problem detected. There are differences in sensitivity and specificity depending on whether patients were admitted for MI, angioplasty, or bypass surgery, with UCDS performing significantly better in MI patients than the other two systems. Positive predictive values were similar for all three methods tested and were only about 25% in this population with a prevalence of 18% quality problems. All systems

could be streamlined for these populations of patients to include only the highest yield screens with little loss of sensitivity.

BACKGROUND

Regulatory efforts to monitor quality of care for payers, like the Health Care Financing Administrations's (HCFA's) contracts with peer review organizations (PROs) are moving toward methods of screening for quality problems that depend on aggregate analyses of processes or outcomes of care for cohorts of patients. For some conditions with adequate numbers of patients per provider or per hospital, such review can be based upon risk-adjusted outcomes assessment. For other high volume patient cohorts, the effort and expense are justified to codify guidelines for care and update them regularly, and for such patients quality of care is likely to be monitored using review against explicit criteria based on these guidelines. However, for most episodes of care, implicit review by physicians of the medical record or of observed medical events, remains, and probably will remain, an important, practical way to identify the rate and type of quality problems that have occurred. Although peer review by one or two physicians is not reliable enough to draw conclusions about an individual case, it is a sound method for identifying patterns and rates of quality problems for groups of patients

The HCFA Health Standards and Quality Bureau contracts with PROs to review medical records of hospitalizations to identify patterns of quality problems experienced by patients treated by a given physician or hospital. Hospitals' internal efforts to assess and improve the clinical quality of care also depend heavily on peer review by physicians. Because physician review is time-intensive, screening systems to identify records with possible quality of care problems have been evolved in an effort to save physician review time by directing it to a higher yield group of records. HCFA first attempted to give guidance in the screening process by disseminating to PROs a set of generic screens. These mainly consisted of a set of adverse outcomes. A similar set of adverse outcome screens was also developed by investigators conducting the Harvard Medical Practice Study, who also required a way to screen records to determine which required physician review. More recently, HCFA developed a record abstraction system known as the Uniform Clinical Data Set (UCDS). Although the UCDS has many other purposes, including collecting data for risk-adjusted outcomes analyses as well as explicit review for patterns of care, early prototypes of the UCDS software included a set of embedded algorithms designed to flag medical records that should be reviewed by a physician.

The performance of screening systems, in terms of sensitivity, specificity, predictive value and cost for identifying quality problems, has rarely been evaluated. Studies have demonstrated the HCFA generic screens and the methods used by one PRO have low predictive value and relatively low sensitivity (Sanazaro and Mills, 1991, Rubin 1992). It was unknown if the set of algorithms developed for UCDS had higher sensitivity and predictive value and how much better performance was achieved given what appeared to be a very extensive data extraction from the medical record. In addition, the cost per quality problem identified had not been determined for any of the screening methods above.

PURPOSE AND AIMS

The purpose of this study was to assess the sensitivity, specificity, predictive value and relative costs of three different systems designed to screen medical records for problems with the quality of hospital care. The systems were 1) a set of screens used by Johns Hopkins Hospital that included HCFA generic screens used by the PROs, 2) the Harvard Medical Practice Study (HMPS) screens, and 3) screening algorithms embedded in version 3.01 of the Uniform Clinical Data Set (hereafter referred to as UCDS). The reference standards used to establish a quality problem were a judgment of substandard ("below standard") care rendered at the end of a structured implicit review method developed at RAND (Rubin et al. 1990,1992), a standardized summary quality score obtained by summing specific aspects of care rated in the structured form, and a judgment of whether or not there was any "quality problem", determined after the rest of the review was complete. We also used a modification of the physician form used to determine negligence used by the Harvard Medical Practice Study (Brennan et al. 1989), but this applied only to records with adverse events and thus could not be used as a general reference standard.

Prior to undertaking the study, we met with Health Standards and Quality Bureau staff at HCFA to determine in which population it might be most helpful to test different screening systems. Staff at that point in time were trying to refine UCDS for various purposes in patients with cardiac disease, and foresaw its evolution into condition-specific data abstraction modules. They believed the results would vary from condition to condition and suggested beginning with cardiovascular disease as HCFA was beginning with those conditions in its efforts to evolve other forms of pattern and outcome analysis as well Therefore, this study was limited to patients with coronary artery disease, specifically, those entering the hospital because of a myocardial infarction or to undergo coronary angioplasty or bypass surgery.

Subsidiary aims of the study were to measure performance of the screening systems at different levels of comorbidity and severity of illness at admission, as well as to measure the interrater reliability of screening methods and physician review as applied in a hospital user's quality assurance program.

METHODS

Overview

A sample of medical records, stratified by whether or not they had passed the hospital's own screening review (which included the HCFA generic screens) and also stratified by whether they were admitted for MI, PTCA or CABG, were screened by trained nurses using all three systems. Each medical record also underwent two independent reviews by physicians. Interrater agreement was assessed in a subsample of cases abstracted by two nurses, and for the total sample between the two independent physician reviews. APACHE III and Charlson scores were calculated for each record as well.

Selection of Records

Records were from patients admitted to Johns Hopkins Hospital, an 1100-bed urban teaching hospital. The hospital is a referral center for PTCA and CABG from Baltimore, other counties in Maryland, and the Washington, D.C. area; approximately 1200 PTCA and 800 CABG operations are performed annually. Routinely, quality assurance nurses review 100% of records using a set of generic indicators that included HCFA's generic screens as a full set until mid-1991. Records were sampled from 7/1/89 through 6/30/91. CABG cases were defined as ICD-9-CM codes 36.10-36.19; PTCA cases as ICD-9-CM codes 36.01,36.02, 36.05; and MI cases as ICD-9-CM diagnoses codes 410.X1.

The sample within each cohort was also stratified by whether or not the record had failed the hospital's own quality assurance screen, in order to enrich the sample with quality problems. The size of the sample was determined by how large a sample would be needed to estimate a true sensitivity of 50% to within 20% (other true values of sensitivity would require a smaller sample size). The sample sizes needed were:

	True population	Needed Sample	Sample
CABG/Fail	281	72	177
CABG/Pass	1281	89	36
PTCA/Fail	136	57	114
PTCA/Pass	1169	89	32
MI/Fail	41	29	34
MI/Pass	243	69	56
Total	3151	405	460

Fifty records were re-abstracted to give us 80% power to detect at an alpha error of .05 the difference between a reliability for two groups of 40% and 60%, or a difference between reliability for two groups of 60% and 75%.

Record Screening

One of the physician investigators (L. Camacho) underwent training at Wisconsin PRO in UCDS version 3.01 record abstraction. He trained quality assurance nurses at Johns Hopkins to abstract records using UCDS in the medical records department of the hospital, using laptop computers. Data elements collected by UCDS include sociodemographics, admitting and discharge information, medical history and physical examination, laboratory and diagnostic tests, endoscopy, operative episodes, treatment interventions, and hospital course (adverse events, admission to special units, etc.) Using automated algorithms, UCDS then generates a summary of whether the case has failed any of its flags. The UCDS algorithms are organized into generic screens, elective surgery, disease-specific screens, organ-specific screens, and discharge status. In each decision area, algorithms look for a primary condition which defines the cases and the context in which the screen is applicable. Then it looks for a secondary condition that triggers a specific flag. For example, if a blood transfusion was given (primary condition) and the patient had an unexpected event of shock due to drug or transfusion reaction within one day after the blood transfusion (secondary condition), the following flag is set: "Shock due to transfusion reaction within one day of blood administration". A complete listing of UCDS flags is available in accompanying documentation (Camacho 1994).

The same nurses applied Johns Hopkins Hospital and Harvard Medical Practice Study screens, which had also been automated on the same computers. Johns Hopkins Hospital screens included HCFA's set of generic screens. They are events or situations thought to signal a potential for poor quality care. For example, cases of unscheduled return to surgery on the same hospital admission for the same condition as the previous surgery or to correct operative problems, would fail screening. The full set of JHH screens are described elsewhere (Camacho 1994, Appendix B) Harvard Medical Practice Study screens are sixteen events associated with potential poor quality care, overlapping somewhat with JHH screens (Camacho 1994, Appendix B).

Quality control for nurse abstraction: All nurse abstractions were overread by one of the investigators to confirm accuracy of identification variables. Dr. Camacho abstracted data for comorbidity and severity of illness scores, and in the process became acquainted with each record well enough to detect if inappropriate flags had been triggered in UCDS due to inaccurate data abstraction. Each UCDS printout was overread and the medical record checked to confirm aberrant lab data or diagnostic tests, or suspicous-looking data.

Physician Review

Each record was reviewed by two or three physicians using 1) a modification of a structured implicit review form for hospital care for diverse medical and surgical conditions (Rubin et al. 1990). Modifications consisted of adding items asking for brief descriptions of any quality problems and justifications for some answers, adding tests and treatments relevant to cardiac care, and excluding some items that had not performed well in other settings; and 2) a condensed form of the adverse event analysis form from the Harvard Medical Practice Study (Brennan et al. 1989), eliminating items asking about the organ system involved in the adverse event, where it took place, the responsible specialty, and others that were not relevant to this study. Items were retained that asked about whether there was injury (morbidity not caused by the disease process), adverse events (injuries caused by medical management), negligence (failure to meet the average practitioner's standard of care at the time care was given), and scores for the confidence in medical causation of the adverse event and confidence that negligence occurred. We defined negligence or adverse event as requiring a score of 4 or higher on the 6-point scale. At the end, physicians were asked on a "physician opinion form" to review the screening results (previously sealed) to comment on whether any quality problems found were related to the screens the case had failed, and invited a critique of the screens as they worked in this particular case.

Determining reference standard from physician review

To increase the reliability of reference standards, we defined "substandard" care or quality problem only when two reviewers agreed upon the determination. For the summary score, we used the mean of the two reviews.

The summary score was calculated using only items on the structured implicit review form that showed some predictive ability for selected outcomes (injury, adverse events, and disability) in a multiple regression model. The final model had 14 variables to which weights were assigned according to the strength of their relationship to disability for that patient cohort.. To combine items with different scales, items were scored as a ratio of assigned to applicable points (RAAP). Summary scores were also adjusted for reviewers' own mean by substracting the RAAP from that reviewer's mean for all cases he had reviewed, and dividing by the standard deviation for the whole sample. A cutoff standardized score was then determined that maximized agreement with determinations of substandard care, quality of care problems, and negligence.

In this report, we emphasize findings using the reference standard of a "quality problem" from the physician's opinion form, as this definition identifies opportunities for improvement even when the care on average was considered "standard" or "above standard" on the implicit review form. Analyses using substandard care and the summary score as the reference standard, as well as analyses using a quality problem judged by either reviewer rather than both, are reported in detail elsewhere (Camacho 1994, Appendix D).

Assessment of reliability

A detailed description of the methods used to assess reliability is available (Camacho 1994). In brief, we calculated crude ageement, kappa, and two alternatives to kappa that have been recently described: the achievement ratio, and kappa-free. These two statistics attempt to adjust the kappa calculation for the maximum kappa achievable using a particular distribution, as kappa is misleading when events are distributed in a highly skewed fashion, such as quality problems. Kappa-free assumes that the a priori probability of a quality problem is 50% in a reviewers mind, rather than the distribution found post-hoc.

In all the reported analyses, we assess the reliability of one reviewer's judgment. However, as a reference standard in the study, we used judgments agreed upon by two reviewers which greatly increases the reliability of the findings. We also calculated the estimated kappa of such two-reviewer consensus judgments but did not make these calculations using kappa-free and the achievement ratio.

Peformance of screening systems

Performance of screening systems against the reference standard was assessed using sensitivity and specificity, predictive value, and the number of nurse-hours and physician-hours per quality problem detected. For the latter, we assumed half an hour per record for physician review. For screening systems, we used HCFA's estimate of 65 minutes for UCDS 3.01,and 15 minutes for JHH or HMPS based on our own experience and the quality assurance nurses' experience. Performance of the screening system was compared to the sensitivity, specificity, predictive value and nurse and physician hours per quality problem if the same number of records were selected at random rather than screened in order to be referred to physicians.

All results, both of analyses of interrater agreement and of screening system performance, pool the results of the two strata within each condition using a reweighting that reflects the original proportions of each stratum for that condition as detailed in the sample selection above. When results are pooled across conditions, they are also reweighted to reflect the proportions of admissions with that condition in the hospital during the time period examined in the study.

RESULTS

Sample

About 65% of records were available for review; thus additional records were requested at random until the appropriate sample size was reached. A total of 460 records were examined. There were 511 nurse abstractions and 951 physician reviews. Nine cases were excluded because the principal procedure was wrongly coded, the clinical condition of interest was not present, or there were too many missing parts to allow a review. 28 cases were excluded because one of the physician reviews did not occur. Fourteen percent of cases were reviewed by a third reviewer.

CABG patients were predominantly white males in their sixties or seventies, retired, covered by private health insurance and/or Medicare, who had had a previous hospital admission for a related condition in the prior year. Coronary atherosclerosis was the most common principal diagnosis, they were discharged to home in two weeks, on average.

PTCA patients were typically white, slightly more likely to be male, between 55 and 70, employed or retired and capable of self-care. They were covered by private insurance, Medicare, or both, and often had not been hospitalized previously for a related condition. Their principal diagnosis was most likely to be coronary atherosclerosis, or, less often, an MI, and they were less severely ill with less comorbid disease than CABG and MI patients. A substantial proportion had adverse occurrences, but few had intra-operative adverse events (PTCA is defined as an operation by UCDS). Discharge to home occurred typically in 5 days in this time period between 1989 and 1991.

In the MI stratum, patients were typically white or black males, aged 40 to 80. Most were retired or employed, and were covered by Medicare or private insurance. Most were capable of self-care before the MI, had not been hospitalized for a related condition, but were sicker at admission and had more comorbid conditions than CABG and PTCA patients.

These demographic and clinical characteristics of the sample are detailed extensively elsewhere (Camacho 1994, pp. 98-120).

Screening Results

UCDS adverse occurrences were more frequent in the CABG stratum (50%) than in PTCA (30%) or MI (30%) strata. Pulmonary edema, cardiac arrest, sustained ventricular tachcardia, atelectasis, pneumonia, and urinary retention were the most common adverse occurrences (Tables 1 and 2). Four patients had an MI after admission which became their principal diagnosis, but were also counted as adverse occurrences. The rate of adverse occurrences per 100 patient-days was 5.8 for CABG, 5.7 for PTCA and 5.0 for MI. The length of hospital stay increased with increasing numbers of adverse occurrences.

Trauma in the hospital was concentrated in the CABG group (Table 3), and was rare in the other strata. Intra-operative adverse events are listed separately in Table 4 We did not include intraoperative blood loss as an adverse event, as UCDS defines it as any blood loss more than 50 cc during an operative procedure. The most frequent intraoperative adverse occurrences were sustained arrhythmia, myocardial infarction and hypotension. Intraoperative adverse occurrences in UCDS do not all raise flags for quality problems. Additional criteria were needed to raise a flag. Adverse occurrences in UCDS should be distinguished from "adverse events" as defined in the Harvard Medical Practice Study, meaning that the event be caused by medical management. UCDS adverse occurrences are events regardless of causation.

Physician Review: Interrater Agreement and Results

One physician reviewed 82% of the cases. Therefore, agreement measurements obtained often reflect agreement of that reviewer with the other seven reviewers. Chance-corrected agreement ranged from slight to moderate when examining this reviewer's agreement with other reveiwers. A detailed analysis of individual reviewers' tendencies is detailed elsewhere (Camacho 1994).

These reliability assessments are for the reliability of one physician's review. Because of poor reliability, we used as the reference standard for the study a quality problem that both physicians agreed about. The estimated reliability for this is more than twice as high as the reliability statistics presented here for interrater agreement.

Table 5 indicates agreement about substandard care, which was 79%, with kappa 0.11, and kappa-free 0.57. Only 2.8% of cases had below standard care after re-weighting for the stratification by cases that failed JHH screens; 2.5% for CABG, 3.2% for PTCA and 2.0% for MI. Tables 6-8 detail agreement about substandard care for each condition.

Table 9 indicates that agreement about a "quality problem" was 68%, with kappa 0.29, the achievement ratio (kappa/maximum achievable kappa) 0.33, and kappa-free 0.37. Of all cases pooled across all strata, 17.8% had a quality problem agreed upon by two reviewers, while 49.5% had a quality problem determined by at least one reviewer.. Proportions of quality problems for the different conditions ranged from 15% to 34% in single reviews, and from 4% to 18% for ratings agreed upon by two reviewers (Tables 10-12). Crude agreement ranged from 67% to 76%, with CABG and MI having poor agreement using the traditional kappa measure. Kappa-free is a better statistic for this highly skewed data, and was 0.33. 0.38, and 0.51 for CABG, PTCA, and MI respectively.

Quality problems and substandard care were closely associated (Table 13). A quality problem can be more minor, such that care overall is still judged standard or above. Thus a quality problem agreed on by two reviewers occurred in 15 of 15 records that both reviewers agreed had substandard care, but also in 24 of 359 records that both reviewers agreed had care that was standard or above.

Nurse Abstractor Agreement

We examined agreement based on nurse-abstracted data, for UCDS, and based on nurse application of screens, for JHH and HMPS, about whether or not a case failed a screen and needed referral for physician review.

Agreement about referral was 98% for UCDS (Table 14), while agreement using each subset of UCDS flags (the five different types of flags included) ranged from 52% to 98%. At an individual screen level agreement was nearly perfect usually exceeding 95%.

For example, agreement about an unexpected event following surgery as determined in UCDS was 98% with kappa .95, kappa-free 0.96 (Table 16). Agreement about inappropriate hospitalization and services for ischemic heart disease and chest pain in UCDS was 96% with kappa .65, kappa-free 0.92 (Table 17).

Agreement for the complete set of JHH screens was 76% (Table 17). However, for most individual screens, chance-corrected agreement was substantial to almost perfect (Tables 18-19). Only for one screen was agreement low (Table 20).

Table 21 shows results for the HMPS set of screens as a whole. Again, agreement was higher for individual screens (Tables 22-23).

Screening Results

Proportions of screen failures varied from 68% to 38% across conditions and screening systems (Table 24). A large proportion of CABG patients failed JHH and HMPS screens, while in UCDS the proportion was lower. For PTCA and MI the pattern was the opposite. Pooling all, the difference in the proportion of screen failures was similar. Cases that passed UCDS but failed JHH and HMPS did so because of "hospital-incurred trauma" or "other undesirable outcomes", "complication of diagnostic or invasive procedure", "abnormal diagnostic finding not addressed", and a few other screens. For all screening systems, patients who failed had higher lengths of stay, more comorbidity and higher severity of illness.

Table 25 details frequency of UCDS flags by main decision areas. A high proportion of cases were flagged for iatrogenic events, failure to establish severity of illness, insufficiency of acute care services, and unexpected event following surgery. The most frequent flags for CABG were procedure complications, unexpected events within three days of surgery, severity of illness criteria not established for calcium abnormalities and acid-base disturbances, and acute care services not rendered for anemia. All the same flags except acid-base disturbances were frequent for PTCA. Lack of follow-up of chest-x ray abnormalities and severity of illness not established for pulmonary abnormalities occurred in a substantial proportion of cases. Again, anemia, calcium abnormalities, pulmonary abnormalities, and lack of follow-up for chest x ray abnormalities were among the most frequent for MI, as well as unexpected death. Extensive details are given elsewhere on the frequency of inidividual UCDS, JHH and HMPS flags (Camacho 1994, Appendix D).

The most frequent JHH screens failed were abnormal diagnostic findings not addressed, hospital-acquired infections, and complications of diagnostic or invasive procedures.

The most frequent HMPS screens failed were for subsequent admissions for the same condition, hospital-incurred trauma, cardiac-respiratory arrest (CABG and MI), death (CABG and MI) and "other".

Screening System Performance

Table 26 indicates the performance of the three screening systems in each of the three conditions and for the records pooled, reweighted according to the representation of these conditions at Johns Hopkins. UCDS was most sensitive compared to the other screening systems for PTCA and MI, although confidence intervals overlap widely. JHH screens were more sensitive for CABG. The "pass" stratum in the MI cohort had no quality problems so that a pooled sensitivity is misleading; there was no variance between the strata). The sensitivity was high for the other stratum which had quality problems.

The specificity was lower, less than 60% in most cases. Predictive values were similar across screening systems except in the MI stratum, which had a small number of confirmed quality problems. For most screening systems only 25% of cases failing screens had confirmed quality problems. This is not much better than the 18% of cases that would have confirmed quality problems if records were selected for review at random.

On average, it took 6.7 hours of record abstraction with UCDS to detect one quality problem, and 2 hours of nurse screening with JHH or HMPS criteria.

Performance of Screen Subsets

The contribution of each screen was assessed by deleting it from the entire set and comparing the performance of the remaining subset with the complete set. Few flags had a measurable effect on the performance of the UCDS system as a whole (Table 27). Six of the UCDS flags were two or more times more likely to be found in cases with a quality problem, whereas four were more likely to be found in cases without a quality problem. The effect of some screens varied across conditions, and these are detailed elsewhere (Camacho 1994, Appendix D, Tables D23 to D25).. Withdrawing flags serially (Figure 1) caused UCDS screening to become less sensitive and more specific as expected. Two flags were very important to maintaining the sensitivity of UCDS to quality problems for these conditions. These were "diagnosis codes include complications of procedures" and "acute care services not rendered for anemia". Other flags caused only small changes in accuracy. Manipulating the screen set did not improve the positive predictive value much.

For JHH screens, five of the screens showed measurable effects on the efficiency and performance of the JHH system (Screens 1025, 154, 101, 155, 156, see Table 28). The most frequent flags are also the ones with the lowest yield and low predictive value with likelihood ratio close to 1. Screens 155 and 156, "IV fluids after midnight on day of discharge" and "purulent or blood draining or wound or open area" occurred in only 1.5% of cases but had high likelihood ratios. Three screens, "reintubation within 72 hours of extubation", "patient discontinuous of lines while in the SICU", and "hospital-acquired decubitus ulcer" were found more often in cases without a quality problem (likelihood ratio < 1). Noteworthy is the high performance of screen 109, "care or lack of care resulting in serious complications" which was a very subjective screen applied by quality assurance nurses. For JHH screens, the plot portraying changes as screens were serially withdrawn (Figure 2) showed a decrease in sensitivity and increase in specificity with each point representing a subset without one or more screens. Maximum accuracy was obtained with the complete set of screens. However, removing the screens with poor performance did not improve the accuracy of the entire system very much. Details of this analysis can be found elsewhere (Camacho 1994, Appendix D, Table D29)

For HMPS, three screens (2,3, and 15) showed substantial impact when excluded from the whole set of HMPS screens (Table 29). Most important were "subsequent admission related to index admission", and "other undesirable outcome". "Hospital-incurred trauma not resolved, or extending hospitalization", also had a small adverse effect on sensitivity. Again, six screens occurred more frequently in cases without a quality problem. As with other screening systems, the plot of sensitivity as a function of the false-positive rate did not show substantial improvement of the subsets with the remaining screens (Figure 3). The likelihood ratio peaks at 1.7 after removing the screens with the lowest predictive value (Camacho 1994, Appendix D, Table D-32).

Influence of Comorbidity and Severity

Table 30 illustrates that screening results differed for different levels of severity of illness, but sample sizes were too small to place much confidence in these results. In addition, the associations with severity and comorbidity differed by condition. In general, all screening systems had higher sensitivity in more severe cases, and higher specificity in more severe cases. There were only minor differences in predictive value with level of severity. As far as comorbidity, HMPS and JHH systems were more sensitive in cases with higher comorbidity, whereas UCDS sensitivity was not affected by comorbidity. For all screening systems, specificity was higher when comorbidity was lower. Predictive value of JHH and HMPS were higher among cases with lower comorbidity, while for UCDS differences were very small.

Performance of screening systems and screens expressed using adjusted odds ratios

Adjusting for comorbidity and severity, one can determine the odds ratios for cases that failed each screening system relative to those that passed (Table 31). Quality problems were 4.8 times more likely among records that failed UCDS than those that passed (Table 29). The equivalent numbers are 2.6 for JHH screening, and 2.5 for HMPS. Because of the sample size, these latter two cannot be distinguished from 1.

Tables 32-34 give adjusted odds ratios for individual UCDS, JHH and HMPS screens. The subsets of screens that gave the best prediction of quality problems had eight screens for UCDS (Table 32). Two of these screens actually had a negative association with a quality problem, "unexpected inpatient event within 3 days of surgery or endoscopy" and "decubitus ulcer acquired in the hospital". Five of the screens belong to the "iatrogenic events" module of UCDS algorithms. To be clear on what this means, it means that system performance is best when records with these screens are actually removed from those referred for physician review regardless of whether they are flagged for any other reason.

Six JHH screens showed significant associations with quality problems (Table 35). Three were related to patient stability at discharge, one flagged unexpected deaths, another was related to medication error or drug transfusion reaction, and one was subjective impression of a quality problem by the nurse.

Among the five HMPS screens comprising the best set selected (Table 36) , three had a negative association with a quality problem ("unplanned return to the OR", "hospital acquired neurological deficit", "cardiac or respiratory arrest").

CONCLUSIONS AND RECOMMENDATIONS

Reliability of Screening

Nurse-applied screening and data abstraction was highly reliable as applied in this hospital, especially with respect to individual screens but also with respect to the overall decision about whether to refer a record to a physician for review.

Reliability of Physician Review

As reported in many previous studies, individual record review by physicians was unreliable. Reliability in this study was even lower than previously achieved by Rubenstein et al. (1990) or Rubin et al. (1992) using a similar review tool, perhaps due to less training time of reviewers and the relative inexperience of the cardiology reviewers, who had never performed peer review before, unlike previous studies that utilized some seasoned reviewers. At the level of reliability achieved, we believe that even for group comparisons such as in this study, a consensus judgment of two reviewers should be used as a reference standard given the low reliability of the individual physician's judgment.

Quality Problems vs. Substandard Care

We confirmed that records judged as substandard by reviewers are a small subset of those in which a quality problem occur. When the technology of structured peer review is used for quality improvement, as we believe it is best used, more opportunities may be identified if reviewers are asked to identify quality problems rather than substandard care. Due to this finding, we changed from our intended reference standard of substandard care to a reference standard of a "quality problem" for this report.

Performance of Outcome-based Generic Screens

A few outcome-based generic screens perform surprisingly well and this subset of screens may be valuable to use for cardiac patients. For the two systems as a whole evaluated here (JHH and HMPS), the predictive value was low (25%) and the sensitivity in the range of 60%. It cost about 2 hours of nurse review and 2 hours of physician review per quality problem detected. Several screens could be eliminated that do not significantly predict quality problems, This would reduce the nurse review time slightly, although not substantially as many of the screens that worked well were not specific events but rather global judgments by nurses covering the entire hospital stay that require a complete record review by the nurse. For example, predictive screens in the JHH system included "medication error" and the nurse's subjective impression of a quality problem, and predictive HMPS screens included "other undesirable outcome". Screens that predict the lack of a quality problem that are currently being used in the opposite way in both systems should be eliminated. Our findings are based only on one hospital, and should be confirmed in other settings.

A good argument can be made for random selection of records for physician review. At the prevalence of quality problems in this particular study, with random selection of records, there would be no nurse hours and from two and a half to three hours of physician review per quality problem detected. Depending on the relative salaries of nurses and physicians in a hospital, it may be cost-effective compared to generic screening to select records at random in order to find the same number of quality problems.

Performance of UCDS version 3.01 screens

While UCDS has many other purposes, including the collection of data for patterns of process use and appropriateness and comparisons of risk-adjusted outcome in different centers, our study evaluated the usefulness of the UCDS algorithms as screening tools for quality problems at the level of the hospital user. For this purpose, UCDS was highly sensitive to quality problems in the cohorts studied but had the same poor predictive value as generic outcome screening systems, referring many records with no quality problems. UCDS had much higher cost per quality problem detected. From the hospital administrators point of view, UCDS certainly would not be cost-effective for identifying quality improvement opportunities as a screening tool for peer review. From a PRO point of view, the additional cost of UCDS also does not justify its use, as 100% record review by physicians could be conducted at the same cost as UCDS abstraction and subsequent physician review of flagged cases. However, UCDS may be cost-effective if used to report outcome or process patterns compared to other hospitals. This remains to be evaluated in other studies by HCFA.

We advocate that our study findings be confirmed in other settings, and used to eliminate ineffective algorithms and those that have negative predictive value for quality problems from the UCDS system as employed for cardiac patients.

A complete listing of quality problems detected is attached (see Appendix A).

REFERENCES

Brennan TA, Localio RJ and Laird NL. Reliability and validity of judgments concerning adverse events suffered by hospitalized patients. Medical Care 1989; 27(12):1148-1158.

Camacho LAB. Comparison of the validity of several symptoms for hospital quality screening and assessment. DISSERTATION submitted to The Johns Hopkins University School of Hygiene and Public Health, 1994.

Rubenstein LV, Kahn KL, Reinisch EJ, et al. Changes in quality of care for five diseases measured by implicit review, 1981 to 1986. JAMA 1990; 264(15):1974-1979.

Rubin HR, Kahn KL, Rubenstein LV and Sherwood MJ. Guidelines for structured implicit review of the quality of hospital care for diverse medical and surgical conditions. Santa Monica, California. A RAND Note, -3006-HCFA. 1990.

Rubin HR, Rogers WH, Kahn KL, Rubenstein LV and Brook R. Watching the doctor-watchers. How well do peer review organization methods detect hospital care quality problems? JAMA 1992; 267(17):2349-2354.

Sanazaro PJ and Mills DH. A critique of the use of generic screening in quality assessment. JAMA 1991; 265(15):1977-1981.

Table 1

Percent distribution within condition of UCDS' adverse occurrences[a] experienced during the hospitalization. Quality of Care Screening Study. Baltimore MD, 1992/1993

Adverse Occurrence	CABG N=215	MI N=90	PTCA N=146	All Conditions (% ± 1.96 S.E.)
Deep Vein Thrombosis	0.4	0	0.3	0.3±0.1
Pulmonary Embolism	0.2	1.5	0	0.2±0.2
Wound Dehiscence	0.4	0	0.1	0.2±0.1
Hypovolemic Shock	0.1	0.4	0	0.1±0.1
Cardiogenic Shock	0.7	0	0.2	0.4±0.2
Septicemic Shock	0.9	0.4	0.5	0.7±0.2
Shock due to Drug/ Transfusion Reaction	0.2	0	0	0.1±0.1
Cardiac Arrest	5.2	11.1	0.8	3.9±2.1
Hemorrhage Requiring Drainage	4.1	0.4	0.5	2.2±2.1
Pulmonary Edema	5.5	6.2	6.5	6.0±4.2
Sustained Ventricular Tachycardia	3.5	3.2	3.2	3.3±3.0
Wound Infection/Abscess	1.7	1.5	0.3	1.1±0.3
Myocardial Infarction	1.0	2.8	1.1	1.2±1.0
Cerebrovascular Accident	1.8	0	0.2	1.0±0.2
Respiratory Failure	1.5	2.4	0.1	1.0±0.3
Urinary Retention	4.8	0	0.2	2.5±0.3
Atelectasis	5.2	0	0.2	2.7±2.8
Pneumonia	6.1	3.9	0.3	3.5±2.9
Other	33.9	11.9	19.8	26.1±7.4

[a] Intra-operative adverse occurrences are presented separately.

Table 2

Percent distribution within condition of the number of UCDS' adverse occurrences[a] experienced during the hospitalization. Quality of Care Screening Study. Baltimore MD, 1992/1993

# of Adverse Occurrences	CABG N=215	MI N=90	PTCA N=146	All Conditions (% ± 1.96 S.E.) N=451
0	50.9	70.2	70.8	62.6±8.0
1	34.2	19.0	16.8	25.6±7.4
2	9.7	6.2	7.9	8.6±4.2
3	1.2	3.7	0.4	1.1±0.3
4	0.8	0.8	0	0.5±0.2
> 5	3.1	0	0.1	1.6 ±2.0

$\chi^2_{(4)} = 25.039$; $p < 0.000$

[a] Intra-operative adverse occurrences are not included.

Table 3

Percentage of cases within condition with trauma (UCDS database) suffered in hospital

. Quality of Care Screening Study. Baltimore MD, 1992/1993.

Type of Trauma	CABG N=215	MI N=90	PTCA N=146	All Conditions (% ±1.96 SE.) N=451
Fall or Accidental Injury	4.9	0	0.1	2.5±2.8
Decubitus Ulcer (In Hospital)	2.5	0.4	0.1	1.3±0.3
Control of Function Loss	0.2	0	0	1.0±0.1
Injury with Etiology Unknown	0.1	0	0	0.0±0.0

Table 4

Percentage of intra-operative adverse occurrences among operative procedures performed (UCDS database), within condition. Quality of Care Screening Study. Baltimore MD, 1992/1993

Adverse Occurrence	CABG N*=307	MI N*=11	PTCA [a] N*=172	All Conditions N*=500
Hypotension	0.60	0	3.00	1.54
Myocardial Infarction	0.07	0	3.40	1.45
Seizure	0	0	0.07	0.03
Sustained Arrhythmia	4.94	0	0.69	2.74
Cardiac Arrest	0.74	0	0.56	0.60
Perforate/Excise	0.13	0	0.35	0.21
Stroke	0.47	0	0	0.23
Return to O.R. [b]	3.15	2.06	0.56	1.98

[a] PTCA is counted as an operative procedure in UCDS database.

[b] Patient needed further surgery as a result of complications arising from a previous operative episode within the same hospitalization. Return to the O.R. is included in the total.

N* : Number of operative episodes.

Table 5

Percentage of ALL cases found to have substandard care in two independent physician reviews of medical records. Quality of Care Screening Study. Baltimore MD. 1992/1993

		Review 1		
		Below Std.	Std.	Total
Review 2	Below Std.	2.8	7.3	10.1
	Std.	14.0	75.9	89.9
	Total	16.8	73.2	100.0

Crude Agreement: 79.0% ; Kappa: 0.11 ± 0.09 ; $\kappa/\kappa_{max} = 0.11 / 0.71 = 0.15$;

$\kappa_{free} = 0.57$; Reliability of substandard care agreed upon by two reviewers: $\kappa = 0.20$

Table 6

Percentage of CABG cases found to have substandard care in two independent physician reviews of medical records. Quality of Care Screening Study. Baltimore MD, 1992/1993

		Review 1		
		Below Std.	Std.	Total
Review 2	Below Std.	2.5	3.5	6.0
	Std.	11.9	82.1	94.0
	Total	14.4	85.6	100.0

Crude Agreement: 84.6% ; Kappa: 0.18 ± 0.12 ; $\kappa/\kappa_{max} = 0.18 / 0.55 = 0.32$; $\kappa_{free} = 0.69$; Reliability of substandard care agreed upon by two reviewers: $\kappa = 0.31$

Table 7

Percentage of PTCA cases found to have substandard care in two independent physician reviews of medical records. Quality of Care Screening Study. Baltimore MD. 1992/1993

		Review 1		
		Below Std.	Std.	Total
Review 2	Below Std.	3.2	12.2	15.4
	Std.	17.8	66.8	84.6
	Total	21.0	89.0	100.0

Crude Agreement: 70.0% ; Kappa: - 0.001 ± 0.16 ; κ/κ_{max} =- 0.001 / 0.813 = 0.123 ; κ_{free} = 0.4 ; Reliability of substandard care agreed upon by two reviewers: κ = - 0.002

Table 8

Percentage of MI cases found to have substandard care in two independent physician reviews of medical records. Quality of Care Screening Study. Baltimore MD. 1992/1993

		Review 1		
		Below Std.	Std.	Total
Review 2	Below Std.	2.0	3.7	5.7
	Std.	7.1	87.2	94.3
	Total	9.1	90.9	100.0

Crude Agreement: 89.3% ; Kappa: 0.23 ± 0.20 ; κ/κ_{max} = 0.23 / 0.75 = 0.31 ; κ_{free} = 0.78 ; Reliability of substandard care agreed upon by two reviewers: κ = 0.37

Table 9

Percentage of cases with a quality problem (QP)[*] according to two reviews of all cases. Quality of Care Screening Study. Baltimore MD, 1992/1993

		Review 1		
		QP	No QP	Total
Review 2	QP	17.8	13.4	31.2
	No QP	18.3	50.5	68.8
	Total	36.1	63.9	100.0

[*] Defined as failure to meet the average practitioner's standard of care.

Crude Agreement: 68.3% ; Kappa: 0.29 ± 0.10 ; $\kappa/\kappa_{max} = 0.29 / 0.89 = 0.33$; $\kappa_{free} = 0.37$; Reliability of quality problem agreed upon by two reviewers: $\kappa = 0.45$

Table 10

Percentage of cases with a quality problem (QP) according to two reviews of CABG cases. Quality of Care Screening Study. Baltimore MD, 1992/1993

		Review 1		
		QP	No QP	Total
Review 2	QP	10.3	12.3	22.6
	No QP	21.0	56.4	77.4
	Total	31.3	68.7	100.0

Crude Agreement: 66.7% ; Kappa: 0.16 ± 0.13 ; $\kappa/\kappa_{max} = 0.16 / 0.78 = 0.21$; $\kappa_{free} = 0.33$; Reliability of quality problem agreed upon by two reviewers: $\kappa = 0.28$

Table 11

Percentage of cases with a quality problem (QP) found in two reviews of PTCA cases. Quality of Care Screening Study. Baltimore MD, 1992/1993

		Review 1		
		QP	No QP	Total
Review 2	QP	18.3	15.2	33.5
	No QP	16.0	50.5	66.5
	Total	34.3	65.7	100.0

Crude Agreement: 68.8% ; Kappa: 0.30 ± 0.16 ; $\kappa/\kappa_{max} = 0.30 / 0.98 = 0.31$; $\kappa_{free} = 0.38$; Reliability of quality problem agreed upon by two reviewers: $\kappa = 0.46$

Table 12

Percentage of cases with a quality problem (QP) according to two reviews of MI cases. Quality of Care Screening Study. Baltimore MD, 1992/1993

		Review 1		
		QP	No QP	Total
Review 2	QP	4.2	10.8	15.0
	No QP	13.5	71.5	85.0
	Total	17.7	82.3	100.0

Crude Agreement: 75.7% ; Kappa: 0.11 ± 0.21; $\kappa/\kappa_{max} = 0.11 / 0.90 = 0.12$; $\kappa_{free} = 0.51$; Reliability of quality problem agreed upon by two reviewers: $\kappa = 0.20$

Table 13

Substandard care vs. quality problem assessed by two physician-reviewers. Quality of Care Screening Study. Baltimore MD, 1992/1993

Results of two reviews combined		Quality Problem			
		Present	Uncertain	Absent	TOTAL
Substandard Care	Present	15	0	0	15
	Uncertain	17	29	3	49
	Absent	24	72	263	359
	TOTAL	56	101	266	423*

Spearman Coefficient Correlation: 0.529 (p=0.037)

* The number differs from the total sample of 460 because nine records were unavailable for review and 28 were reviewed by only one physician.

Table 14

Agreement between Nurse-abstractors. UCDS flags raised by all decision areas. Quality of Care Screening Study. Baltimore MD. 1992/1993

		Abstractor 1			
	Flag	'PA'	'MO'	Absent	TOTAL
Abstractor 2	'PA'	42	0	0	42
	'MO'	0	5	0	5
	Absent	1	0	3	4
	TOTAL	43	5	3	51

'PA' : Refer case to Physician Advisor. 'MO': Possible problems. Monitoring of patterns warranted. Absent: the case passed screening

Crude Agreement: 98.0% ; Kappa: 0.93 ± 0.13 ; $\kappa/\kappa_{max} = 0.93 / 0.93 = 1.0$;

$\kappa_{free} = 0.97$

Table 15

Agreement between Nurse-abstractors. Flag GS05PAI1 (Unexpected event following surgery). Quality of Care Screening Study. Baltimore MD, 1992/1993

		Abstractor 1		
	GS05PAI1	Present	Absent	TOTAL
Abstractor 2	Present	15	0	15
	Absent	1	34	35
	TOTAL	16	34	50

Crude Agreement: 98.0% ; Kappa: 0.95 ± 0.08 ; $\kappa/\kappa_{max} = 0.95 / 0.95 = 1.00$; $\kappa_{free} = 0.96$

Table 16

Agreement between Nurse-abstractors. Flag DP01PR8D (Ischemic Heart Disease/Chest Pain. Appropriate Hospitalization and Services). Quality of Care Screening Study. Baltimore MD, 1992/1993

		Abstractor 1		
	DP01PR8D	Present	Absent	TOTAL
Abstractor 2	Present	47	1	48
	Absent	1	1	2
	TOTAL	48	2	50

Crude Agreement: 96.1% ; Kappa: 0.65 ± 0.27 ; $\kappa/\kappa_{max} = 0.65 / 1.00 = 0.65$; $\kappa_{free} = 0.92$

Table 17

Agreement between Nurse-abstractors. Whole set of JHH Screens. Quality of Care Screening Study. Baltimore MD, 1992/1993

		Abstractor 1		
		Fail	Pass	TOTAL
Abstractor 2	Fail	37	6	43
	Pass	6	2	8
	TOTAL	43	8	51

Crude Agreement: 76.5% ; Kappa: 0.11 ± 0.27 ; $\kappa / \kappa_{max} = 0.11 / 1.00 = 0.11$;

$\kappa_{free} = 0.53$

Table 18

Agreement between Nurse-abstractors. Generic screen: 'Hospital-Acquired Infection. Quality of Care Screening Study. Baltimore MD, 1992/1993

		Abstractor 1		
		Fail	Pass	TOTAL
Abstractor 2	Fail	7	2	9
	Pass	2	40	42
	TOTAL	9	42	51

Crude Agreement: 92.2% ; Kappa: 0.73 ± 0.27 ; κ/κ_{max} = 0.73 / 1.00 = 0.73 ; κ_{free} = 0.84

Table 19

Agreement between Nurse-abstractors. Generic Screen: 'Intravenous fluids or drugs after 12 midnight on day of discharge'. Quality of Care Screening Study. Baltimore MD, 1992/1993

		Abstractor 1		
		Fail	Pass	TOTAL
Abstractor	Fail	3	3	6
2	Pass	1	44	45
	TOTAL	4	47	51

Crude Agreement: 92.2% ; Kappa: 0.56 ± 0.27 ; $\kappa/\kappa_{max} = 0.56 / 0.78 = 0.72$; $\kappa_{free} = 0.84$

Table 20

Agreement between Nurse-abstractors. Generic screen: 'Abnormal diagnostic finding not addressed and resolved where the records do not explain why they are not resolved'. Quality of Care Screening Study. Baltimore MD. 1992/1993

		Abstractor 1			
		Fail	Pass	Unknown[a]	TOTAL
Abstractor 2	Fail	28	16	1	45
	Pass	4	2	0	6
	TOTAL	32	18	1	51

[a] Documentation was considered insufficient. This category was included to compute Kappa.

Crude Agreement: 58.8% ; Kappa: -0.03 ± 0.22 ; $\kappa/\kappa_{max} = 0.03 / 0.37 = 0.42$; $\kappa_{free} = 0.20$

Table 21

Agreement between Nurse-abstractors. HMPS Screening Criteria: Whole set. Quality of Care Screening Study. Baltimore MD, 1992/1993

		Abstractor 1		
		Fail	Pass	TOTAL
Abstractor 2	Fail	39	3	42
	Pass	5	5	10
	TOTAL	44	8	52

Crude Agreement: 84.6% ; Kappa: 0.46 ± 0.27 ; $\kappa/\kappa_{max} = 0.46 / 0.87 = 0.54$; $\kappa_{free} = 0.69$

Table 22

Agreement between Nurse-abstractors. HMPS Screening Criterion:'Death, except (1) in planned terminal care, (2) DNR order, and (3) patient is gravely ill from a major disease process (massive trauma, thermal burns > 50%, etc.)'. Quality of Care Screening Study. Baltimore MD, 1992/1993

		Abstractor 1		
		Fail	Pass	TOTAL
Abstractor 2	Fail	8	2	10
	Pass	1	41	42
	TOTAL	9	43	52

Crude Agreement: 93.8% ; Kappa: 0.81 ± 0.27 ; κ/κ_{max} = 0.81 / 0.94 = 0.86 ; κ_{free} = 0.88

Table 23

Agreement between Nurse-abstractors. HMPS Screening Criterion: 'Cardiac or respiratory arrest, except (1) in planned teminal care, and (2) when patient is gravely ill from a major disease process (massive trauma, thermal burns > 50%, etc.)'. Quality of Care Screening Study. Baltimore MD, 1992/1993

		Abstractor 1		
		Fail	Pass	TOTAL
Abstractor 2	Fail	6	0	6
	Pass	3	43	46
	TOTAL	9	43	52

Crude Agreement: 94.2% ; Kappa: 0.77 ± 0.27 ; κ/κ_{max} = 0.77 / 0.77 = 1.00 ; κ_{free} = 0.88

Table 24

Percentage (±1.96 S.E.) of screen failures in each screening system, within each condition. Quality of Care Screening Study. Baltimore MD, 1992/1993

Screening System	CABG N=215	MI N=90	PTCA N=146	All Conditions N=451
UCDS	54.6±12.9	65.4±9.6	68.0±14.5	61.1±8.8
JHH	67.6±12.4	37.6±8.9	40.4±14.5	53.6±8.6
HMPS	63.0±12.7	46.2±9.7	46.3±15.0	54.6±8.9

Table 25

Frequency [a] of UCDS flags by main decision areas within conditions. Quality of Care Screening Study. Baltimore MD, 1992/1993

Decision Area	CABG n [%]	PTCA n [%]	MI n [%]	All Conditions %
Discharge Status/Disposition	16 [3.7]	12 [2.0]	10 [10.9]	3.3
Medical Stability at Discharge	24 [4.5]	26 [18.6]	11 [11.3]	11.0
Organ Specific A [b]	60 [24.6]	39 [20.7]	30 [30.4]	22.7
Organ Specific B [c]	19 [10.2]	38 [41.4]	42 [48.7]	26.6
Unexpected Event Following Surgery	65 [16.9]	77 [20.6]	3 [2.4]	17.1
Death	37 [5.8]	8 [0.7]	19 [9.2]	4.0
Iatrogenic Events	155 [26.0]	50 [12.7]	12 [6.2]	18.7

[a] Proportions of cases flagged are pooled across strata. Percentages do not add to the total for each condition since many cases have more than one flag. The number of cases is simply the sum of cases flagged disregarding sample fractions. [b] Severity of illness not established. [c] Acute care services not rendered.

Table 26

Performance of Screening Systems in a set of medical records with **17.8 %** found to have a Quality Problem (QP) by two physicians. Quality of Care Screening Study. Baltimore MD, 1992/1993

Screens	% of Records Positive	Sensitivity (% ± 1.96 S.E.)	Specificity (% ± 1.96 S.E.)	PPV[a] (%)	NPV[b] (%)	# hours-MD per QP[c] detected[d]	# hours-nurse per QP detected[e]
UCDS							
ALL	61.6	85.5±6.6	48.3±9.1	26.5± 6.6	93.9± 9.2	1.9	6.7
CABG	53.9	80.6±10.4	51.9±13.0	26.6 [f]	92.5 [f]	1.9	7.0
PTCA	60.1	88.2±9.8	46.0±15.5	26.1 [f]	94.7 [f]	1.9	6.4
MI	67.7	100.0	39.3±11.2	26.3 [f]	100.0 [f]	1.9	5.6
JHH							
ALL	55.4	74.1±12.2	56.8±11.3	26.8±12.8	91.1±10.9	1.9	1.9
CABG	61.6	97.0± 3.8	46.1±17.6	28.1 [f]	98.6 [f]	1.8	1.4
PTCA	36.8	59.7±29.1	68.2±17.0	29.0 [f]	88.6 [f]	1.7	2.3
MI	32.3	14.4 [g]	63.7±11.7	8.0 [f]	77.4 [f]	6.3	9.7
HMPS							
ALL	48.5	69.4±18.4	56.1±11.5	25.6±12.0	89.4±11.4	2.0	2.0
CABG	55.5	79.1±29.1	49.6±17.4	25.5 [f]	91.6 [f]	2.0	1.7
PTCA	42.6	69.7±27.6	63.3±18.0	29.2 [f]	90.6 [f]	1.7	2.0
MI	36.6	14.4 [g]	58.5±12.7	7.0 [f]	75.8 [f]	7.1	9.6

[a] Positive Predictive Value [b] Negative Predictive Value [c] Quality Problem [d] Assuming 1/2 hour per physician review. [e] Assuming 1 hour per nurse abstraction for UCDS, and 15 min. for JHH and HMPS (each). [f] Not a pooled estimate. Calculated from pooled sensitivity, specificity and overall prevalence. [g] No variance within strata.

Table 27

Performance of different subsets of UCDS flags in a set of medical records with 17.8 % found to have a Quality Problem. Quality of Care Screening Study. Baltimore MD, 1992/1993

Flag excluded from set*	% of Records failed screens	Sensitivity (%)	Specificity (%)	PPV [a] (%)	NPV [b] (%)	# hours-MD per QP detected [c]	# hours-Nurse per QP detected[d]
Complete set[e]	57.7	85.5	48.3	26.5	93.9	1.8	6.7
OP05B - Anemia/Polycythemia. Acute care services not rendered.	47.7	60.8	55.1	22.4	86.5	2.2	9.3
GS05I1 - Diagnosis codes include complications of procedures.	48.5	76.3	57.5	28.1	91.2	1.8	7.3
GS02C9 - No follow-up of acutely abnormal X ray.	56.7	81.2	48.7	25.7	92.3	1.9	6.5
OP04A - Leukocytosis/leukopenia. Severity of disease not established.	57.5	85.3	48.5	26.5	93.9	1.9	6.5
OP01A - Endocrine (Calcium). Severity of disease not established.	55.9	85.3	50.6	27.4	94.1	1.8	6.5

continued

Table 27

Flag excluded from set*	% of Records failed screens	Sensitivity (%)	Specificity (%)	PPV [a] (%)	NPV [b] (%)	# hours-MD per QP detected [c]	# hours-Nurse per QP detected[d]
OP07A - Acid-base disturbances. Severity of disease not established.	54.3	84.6	52.3	27.9	94.0	1.8	6.6
OP10A - Pulmonary Disease. Severity of disease not established.	57.6	85.3	48.5	26.6	93.9	1.9	6.7
GS02C24 - IV medication discontinued on day of discharge.	57.3	85.3	48.8	26.7	93.8	1.9	6.6
GS06K5 - Unexpected inpatient event within 3 days of operating room.	55.9	85.3	50.5	27.3	94.1	1.8	6.5
GS06K8 - Cardiac arrest within day of blood transfusion.	57.6	85.5	48.4	26.5	93.8	1.9	6.6
GS06K9 - Death within 1 day of blood transfusion.	57.6	85.5	48.4	26.5	93.8	1.9	6.6
GS06K32 - Accidental puncture of laceration during procedure.	57.6	85.5	48.3	26.5	93.8	1.9	6.6

continued

Table 27

Flag excluded from set*	% of Records failed screens	Sensitivity (%)	Specificity (%)	PPV [a] (%)	NPV [b] (%)	# hours-MD per QP detected [c]	# hours-Nurse per QP detected[d]
"Chance" Perform.[f]	57.7	57.7	42.3	17.8	82.4	2.8	-

* Flag description is truncated. Please, see Table D-11 for full description.

[a] Positive predictive value. [b] Negative predictive value. [c] Assuming 1/2 hour per physician review. [d] Assuming 1 hour per nurse abstraction. [e] The complete set includes screens that contributed with very few cases and were not included in these analyses. [f] No Screening. Proportion taken from complete set.

Table 28

Performance of subsets of JHH Screens formed by exclusion of one screen, in a set of medical records with 17.8 % found to have a Quality Problem. Quality of Care Screening Study. Baltimore MD, 1992/1993

Screen excluded from set*	% of Records failed screens	Sensitivity (%)	Specificity (%)	PPV[a] (%)	NPV[b] (%)	# hours-MD per QP detected [c]	# hours-Nurse per QP detected[d]
Complete Set [e]	48.7	74.1	56.8	27.2	91.0	1.8	1.8
1025 - Complication of diagnostic or invasive procedure.	34.2	57.4	70.9	30.0	88.4	1.7	2.4
154 - Abnormal diagnostic finding not addressed and resolved.	41.7	64.3	63.2	27.6	89.0	1.8	2.2
800 - Hospital-acquired infection.	48.3	74.1	57.3	27.5	91.1	1.8	1.9
125 - Death during or following any surgery.	48.1	74.4	57.4	27.3	90.8	1.8	1.9
101 - Readmission for complication or incomplete hospitalization.	45.7	66.0	58.7	25.8	88.8	1.9	2.1
109 - Care or lack of care resulting in serious complications.	48.6	74.1	57.0	27.3	91.0	1.8	1.9

continued

Table 28

Screen excluded from set*	% of Records failed screens	Sensitivity (%)	Specificity (%)	PPV[a] (%)	NPV[b] (%)	# hours-MD per QP detected [c]	# hours-Nurse per QP detected[d]
803 - Hospital-acquired decubitus ulcer.	48.7	74.1	56.8	27.2	91.0	1.8	1.9
108 - Medication error or adverse drug reaction.	48.3	73.4	57.1	27.2	90.8	1.8	1.9
155 - IV fluids after 12 midnight on day of discharge.	47.8	70.0	57.0	26.2	89.7	1.9	2.0
156 - Purulent or bloody drainage or wound or open area.	47.3	66.0	56.8	25.0	88.5	2.0	2.1
301 - Reintubation of patient within 72 hr. after extubation.	58.7	74.1	56.8	27.2	91.0	1.8	1.9
7500 - Death within 48 hr. of admit through the emergency room.	48.4	74.1	57.2	27.4	91.0	1.8	1.9
"Chance" Perform.[f]	48.7	48.7	51.3	17.9	82.1	2.8	-

* Description of screens is truncated. (Camacho 1994, Table D-12, contains a full description of screens.)

[a] Positive predictive value. [b] Negative predictive value. [c] Assuming 1/2 hour per physician review. [d] Assuming 1 hour per nurse abstraction. [e] Includes flags that contributed with very few cases and are not listed above. [f] No screening. Proportion of failed screens taken from complete set.

Table 29

Performance of different subsets of HMPS Screens in a set of medical records with 17.8 % found to have a Quality Problem. Quality of Care Screening Study. Baltimore MD, 1992/1993

Screen excluded from set*	% of Records failed screens	Sensitivity (%)	Specificity (%)	PPV[a] (%)	NPV[b] (%)	# hours-MD per QP detected [c]	# hours-Nurse per QP detected[d]
Complete Set	48.5	69.4	56.1	25.6	89.4	2.0	2.0
02 - Subsequent admission related to index admission.	36.6	40.7	64.3	19.9	83.3	2.5	3.4
03 - Hospital-incurred trauma not resolved, or extending hospitalization.	43.0	67.5	62.4	28.1	89.8	1.8	2.1
04 - Untoward drug reaction unresolved at discharge, or extending hospitalization.	48.2	69.2	56.3	25.6	89.4	1.9	2.0
05 - Transfer from general care to special care unit.	48.0	69.0	56.5	25.7	89.3	1.9	2.0
06 - Transfer to acute care hospital.	48.5	69.4	56.1	25.6	89.4	2.0	2.0
07 - Unplanned return to operating room.	48.4	69.4	56.2	25.6	89.4	2.0	2.0
08 - Treatment or operation because of damage, or subsequent to invasive proc.	48.7	69.4	56.1	25.6	89.4	2.0	2.0
09 - Acute MI, CVA, or PE during or following surgical or invasive proc.	48.4	69.4	56.1	25.6	89.4	2.0	2.0

continued

Table 29

Screen excluded from set *	% of Records failed screens	Sensitivity (%)	Specificity (%)	PPV[a] (%)	NPV[b] (%)	# hours-MD per QP detected [c]	# hours-Nurse per QP detected[d]
10 - Hospital-acquired neurological deficit.	48.7	69.4	56.1	25.6	89.4	2.0	2.0
11 - Death, except in terminal care, DNR order, gravely ill patient.	46.6	69.2	58.3	26.5	89.7	1.9	2.0
12 - Unexplained abnormal vital sign on last full day of discharge.	46.9	69.4	57.9	26.4	89.6	1.9	2.0
13 - Cardiac arrest, except in terminal care, DNR order, gravely ill patient.	48.3	69.4	56.3	25.7	89.4	1.9	2.0
15 - Other undesirable outcome.	37.6	61.2	67.5	29.1	88.9	1.7	2.3
"Chance" Perform.[e]	48.5	48.7	51.3	17.9	82.1	2.8	-

* Screen description is truncated. (Camacho 1994, Table D-13, contains a full description of screens).

[a] Positive predictive value. [b] Negative predictive value. [c] Assuming 1/2 hour per physician review. [d] Assuming 1 hour per nurse abstraction. [e] No screening. Proportion of failed screens taken from complete set.

Table 30

Sensitivity (%), Specificity (%) and Positive and Negative Predictive Values (%) pooled across conditions, by levels of comorbidity and severity of illness at admission. Mean of estimates from two reviewers. Quality of Care Screening Study. Baltimore MD, 1992/1993

	Severity of Illness		Comorbidity	
UCDS	High	Low	High	Low
Sensitivity	100.0	79.1	87.5	83.6
Specificity	37.8	58.6	43.8	52.4
Pos. Pred. Value [a]	25.9	29.4	25.3	27.7
Neg. Pred. Value [a]	100.0	92.8	94.1	93.1
JHH				
Sensitivity	97.6	60.3	55.2	96.8
Specificity	52.5	62.8	49.0	66.6
Pos. Pred. Value [a]	30.9	26.1	19.1	38.7
Neg. Pred. Value [a]	99.0	87.9	83.4	99.0
HMPS				
Sensitivity	77.6	69.9	65.2	82.7
Specificity	51.9	61.1	48.0	65.2
Pos. Pred. Value [a]	26.0	28.1	21.4	34.1
Neg. Pred. Value [a]	91.4	90.3	86.4	94.5

[a] Calculated based on a common prevalence of 17.8%.

Table 31

Odds ratios (O.R.) for failing UCDS screens for confirmed quality problems: overall and stratified estimates by levels of comorbidity and severity of illness at admission. Quality of Care Screening Study. Baltimore MD, 1992/1993

	Stratum	Stratified O.R.	Overall O.R. (95% C.I.)
Comorbidity [a]	Low	5.00	4.83 (1.18, 19.81)
	High	4.60	
Severity of Illness	Low	4.62	
	High	- [b]	

Tests for interaction: [a] χ^2_1 =0.30 ($p<0.75$) ; [b] Zero cell: UCDS=1 predicts outcome perfectly.

Table 32

Odds ratios (O.R.) for failing JHH screens in cases with a confirmed quality problems: overall and stratified by comorbidity and severity of illness at admission. Quality of Care Screening Study. Baltimore MD, 1992/1993

	Stratum	Stratified O.R.	Overall O.R. (95% C.I.)
Comorbidity [a]	Low	39.87	2.62 (0.83. 8.29)
	High	0.95	
Severity of Illness [b]	Low	1.50	
	High	39.87	

Tests for interaction: [a] $\chi^2_1 = 21.17$ ($p < 0.001$) [b] $\chi^2_1 = 6.94$ ($p < 0.01$)

Table 33

Odds ratios (O.R.) for failing HMPS screens in cases with a confirmed quality problems: overall and stratified by comorbidity and severity of illness at admission. Quality of Care Screening Study. Baltimore MD, 1992/1993

	Stratum	Stratified O.R.	Overall O.R. (95% C.I.)
Comorbidity [a]	Low	6.07	2.54 (0.81, 7.96)
	High	1.44	
Severity of Illness [b]	Low	2.17	
	High	3.54	

Tests for interaction: [a] $\chi^2_1 = 1.35$ ($p<0.25$) [b] $\chi^2_1 = 1.80$ ($p<0.25$)

Table 34

Estimated odds ratios (O.R.) and 95% confidence intervals for individual UCDS screens which best predicted quality problems. Quality of Care Screening Study. Baltimore MD, 1992/1993

Variable	Univariate O.R. (95% C.I.)	Adjusted O. R. (95% C.I.)
Severity of Illness at Admission	0.55 (0.19 - 1.58)	0.28 (0.09 - 0.92)
Comorbidity	1.24 (0.44 - 3.51)	1.45 (0.46 - 4.57)
GS6K32	20.33 (2.48 - 143.52)	28.12 (4.51 -175.12)
GS2C24	5.64 (0.83 - 38.28)	15.14 (1.70 -134.72)
GS6K31	1.72 (0.38 - 7.84)	10.72 (1.62 - 70.96)
GS2C9	4.13 (0.71 - 24.12)	10.72 (1.47 - 78.43)
GS6K8	1.18 (0.29 - 4.79)	8.46 (1.53 - 46.88)
OP05B	2.52 (0.79 - 8.11)	3.71 (1.01 - 13.65)
GS6K5	0.21 (0.06 - 0.73)	0.15 (0.03 - 0.79)
GS6K10	0.19 (0.02 - 1.54)	0.06 (0.00 - 0.96)

Table 35

Estimated odds ratios and 95% confidence intervals for individual JHH screens which best predicted quality problems. Quality of Care Screening Study. Baltimore MD, 1992/1993

Variable	Univariate O.R. (95% C.I.)	Adjusted O.R. (95% C.I.)
Severity of Illness at Admission	0.55 (0.19 - 1.58)	0.07 (0.03 - 0.18)
Comorbidity	1.24 (0.44 - 3.51)	1.71 (0.43 - 6.77)
127	1.66 (0.47 - 5.85)	14.68 (3.15 - 68.32)
155	17.59 (1.96 - 157.78)	232.47 (21.53 - 2509.61)
156	28.67 (3.39 - 242.14)	34.35 (4.56 - 258.48)
151	5.45 (0.39 - 75.92)	28.77 (4.23 - 195.46)
109	6.57 (1.01 - 42.75)	11.64 (2.17 - 62.28)
108	1.04 (0.40 - 2.68)	3.32 (1.04 - 10.63)

Table 36

Estimated odds ratios and 95% confidence intervals for individual HMPS screens which best predicted quality problems. Quality of Care Screening Study. Baltimore MD, 1992/1993

Variable	Univariate O.R. (95% C.I.)	Adjusted O.R. (95% C.I.)
Severity of Illness at Admission	0.55 (0.19 - 1.58)	0.38 (0.11 - 1.34)
Comorbidity	1.24 (0.44 - 3.51)	1.28 (0.40 - 4.11)
02	3.98 (1.16 - 13.62)	4.38 (1.06 - 18.16)
05	5.23 (1.35 - 20.30)	8.02 (1.53 - 42.01)
07	0.12 (0.02 - 0.92)	0.13 (0.02 - 1.00)
10	0.16 (0.02 - 1.28)	0.10 (0.01 - 1.09)
13	0.29 (0.10 - 0.88)	0.29 (0.06 - 1.35)

Table	Appendix page
A1. Listing of adverse events and quality problems by record ID, MI patients (N=90)	1
A2. Listing of adverse events and quality problems by record ID, PTCA patients (N=134)	6
A3. Listing of adverse events and quality problems by record ID, CABG patients (N=213)	14
Abbreviations used in Appendix tables	26

A1. Listing of adverse events and quality problems by record ID, MI patients (N=90)

QID *	Adverse Event	Quality Problem
001	Not specified	None
002	None	Abnormal lab test not addressed
003	Not specified	None
004	Sepsis; GI bleed	None
005	None	None
006	Recurrent ischemia	Heparin not given
007	Cardiac arrest; GI bleed	Failure to install pacemaker or to perform ECG monitoring; failure to perform endoscopy
008	None	On antimicrobial, no culture ordered
009	GI bleed	Unchecked hematocrit fall (on heparin)
	2nd MI after GI bleed; renal failure	Significant hematocrit drop discovered on the next day
010	None	None
011	Postoperative bleeding requiring re-operation	Not described
012	Bleeding from cath site	Abnormal lab value not addressed; premature removal of sheath
013	Worsening renal failure	Delay in sending patient to ICU; premature discharge while creatinine was increasing
014	EMD arrest	Failure to perform cath
015	None	None

QID *	Adverse Event	Quality Problem
016	Groin pseudoaneurysm	None
017	None	None
018	Bacteremia following cath	Not specified
019	None	None
020	None	None
021	None	None
022	None	None
	Transient visual field defect	Heparin discontinued prematurely
023	Recurrent VT arrest while being loaded with amiodarone	Premature discharge from CCU with return 24 hours later
024	Reocclusion of LAD	Inadequate anticoagulation
025	None	None
026	None	None
027	None	None
028	AMI during PTCA attempt	None
029	None	None
030	None	None
031	None	None
032	MI; aspiration pneumonia and sepsis	None
033	None	None
101	None	None
102	Acute GI bleeding; Nosocomial infection (C. difficile colitis)	None

QID *	Adverse Event	Quality Problem
103	None	None
104	None	None
106	Nosocomial infection (UTI)	None
107	None	None
108	None	None
109	None	None
110	None	None
111	None	Prolonged and inappropriate delay in consult and transfer to CCU
112	Acute renal failure	Delay in transfer to CCU
113	None	None
114	None	No risk stratification post-MI with known 70-90% LAD stenosis
115	None	None
116	None	None
117	None	None
118	None	None
119	None	None
120	Stroke	Failure to heparinize fresh MI
121	None	None
122	None	None
123	None	None
124	Gross hematuria from thrombolytic	None

QID *	**Adverse Event**	**Quality Problem**
125	None	None
126	None	None
127	None	None
128	None	None
129	Line sepsis	None
130	None	None
131	None	None
132	None	None
133	None	None
134	Heart block necessitating pacemaker	Poor initial medical treatment
135	None	None
136	None	None
137	None	Abnormal diagnostic finding not addressed
138	None	None
139	Post CABG A-fib	Drug given and blood level not measured; symptomatic bradycardia and no Holter or pacemaker
140	None	None
141	None	None
142	None	None
143	None	None
144	None	None

QID *	**Adverse Event**	**Quality Problem**
145	None	None
146	Hypoglycemia X2	Insulin dose not adjusted at discharge
147	None	Delay in tPA therapy
148	None	None
149	None	None
150	None	None
151	None	None
152	None	None
153	None	None
154	None	None
155	None	None
156	None	None
157	None	None
158	None	None
159	None	Discharge without stress test

* The QID is the record identification number. Records #001-033 had failed screening indicators at JHH; records #101-159 had passed it. Numbering is discontinuous as the leftmost digit identifies the study stratum. Some strata had less than 99 cases. The two rightmost digits of the QID range between 01 and 99; 00 is not used. Forms from record QID #105 were unavailable for review.

A2. Listing of adverse events and quality problems by record ID, PTCA patients (N=134)

QID *	**Adverse Event**	**Quality Problem**
201	Retroperitoneal hemorrhage after groin procedure	None
202	Fall in hematocrit 2nd to groin hematoma	None
203	GI bleeding	No GI prophylaxis despite stress in CCU, thrombolytics, ASA, heparin
204	Groin hematoma	None
205	VT during RCA injection	Over-sedation with Valium, respiratory depression; no ABG
206	Small MI	Unclear
208	None	None
209	Coronary occlusion after PTCA; femoral bleeding after PTCA	None
210	Groin hematoma; GI bleeding	None
211	CHF	Fluid overload from transfusion
212	Groin bleeding, hematoma, requiring blood transfusion, following thrombolytic therapy and cath	None
213	Adverse drug reaction; streptokinase	None
214	Small MI during PTCA	Abnormal diagnostic finding not addressed
215	Pulmonary edema	Poor fluid management

QID *	Adverse Event	Quality Problem
216	Retroperitoneal bleed	Premature discharge; CT to localize pseudoaneurysm not done; hematocrit not checked on discharge
217	Retroperitoneal Bleed	Heparin/PTT not well controlled
218	Coronary occlusion and thrombus during PTCA	Unclear
219	VT/VF arrest during cath	None
220	Ventricular arrhythmias, following acute MI during PTCA attempt	None
221	Acute closure during PTCA; hematoma with falling hematocrit	Abnormal lab finding not addressed (discharged with hematocrit still falling)
222	Ischemia lower extremities following cath, IABP, PTCA	None
223	Groin bleed with hematoma requiring transfusion, and MI during PTCA	Poor evaluation and follow-up of hematoma before discharge
224	None	None
225	DVT following cath/PTCA	Thrombolytic therapy not clearly indicated
226	Coronary reocclusion s/p PTCA	PTCA not indicated
227	Retroperitoneal bleeding; fall in hematocrit	Inadequate evaluation of rapid hematocrit fall
228	Retroperitoneal bleed; CHF post transfusion	Inadequate management of fluid overload
229	Large hematoma with low hematocrit	Unclear
230	Groin hematoma; small MI	None

QID *	Adverse Event	Quality Problem
231	Groin hematoma; femoral nerve palsy	None
232	Renal failure post cath	Poor fluid management
233	Groin pseudoaneurysm	None
234	Bleeding following thrombolytic therapy	None
235	Bleeding post cath	None
236	GI bleeding; allergic reaction to thrombolytic	None
237	Pulmonary edema due to LAD closure; acute renal failure due to dye	None
238	Groin bleed and fall in hematocrit	None
239	Cerebrovascular hemorrhage following thrombolytic	None
240	Groin bleeding; fat necrosis	None
241	Groin hematoma leading to anemia; pseudoaneurysm	None
242	Fall in hematocrit and hematoma; RCA dissection during PTCA	None
243	Coronary dissection during PTCA; groin infection; line infection	None
244	None	None
245	VF arrest during cath; fall in hematocrit during cath.	None
246	Subarachnoid bleed	tPA given after major head trauma and epistaxis

QID *	Adverse Event	Quality Problem
247	Staphylococcus aureus bacteremia with infected hematoma	None
248	VF arrest; sternal wound dehiscence	None
249	Bleeding at cath site requiring transfusion	Unclear
250	Acute renal failure secondary to contrast; thrombophlebitis secondary to IV catheter	None
251	VF arrest during cath	None
252	Groin bleeding and surgery at IABP site	None
253	Dissection and MI during PTCA	None
254	Groin pseudoaneurysm	None
255	Exacerbation of chronic renal failure; MI after cath	None
257	AV fistula after cath	None
258	Acute renal failure, CHF due to IV contrast and volume	Gentamicin given prior to cath; misuse of contrast
259	Total occlusion of RCA following failed PTCA	Consideration of CABG was not documented
260	CHF, UTI, groin bleed	Poor fluid management
261	VF arrest during cath; groin bleed requiring transfusion	None
262	Pulmonary edema during cath	None
263	Line sepsis; deterioration of renal function	Renal function deterioration not addressed

QID *	**Adverse Event**	**Quality Problem**
264	Protamine reaction	None
265	Right femoral artery pseudoaneurysm	Abnormal test result not addressed
266	Cardiac arrest during PTCA	Trauma from guide wire in LV
267	Groin pseudoaneurysm	None
268	RCA dissection and MI post angioplasty	None
269	Acute renal failure	None
270	None	None
271	CVA after cath	None
272	Large groin hematoma	None
273	Groin hematoma	None
274	RCA restenosis	None
275	Acute dissection and MI	None
276	Pseudoaneurysm	None
277	Acute MI due to acute occlusion	Anemia not addressed
278	Line sepsis	None
279	Laceration of right femoral artery; pulmonary edema	None
280	None	None
281	Bradycardia and hypotension	Untreated UTI
282	Bleeding at cath site	Puncture of femoral profunda artery
283	Retroperitoneal bleed; MI s/p angioplasty	None

QID *	**Adverse Event**	**Quality Problem**
284	Line sepsis	Probable break in sterile technique
286	Bacteremia	None
287	Pseudoaneurysm	None
288	MI post PTCA	None
289	None	None
290	Groin hematoma	None
291	MI, 2nd left circumflex artery dissection during PTCA	Discharge planning: antibiotics decided but not prescribed
292	Pulmonary edema and increased CPK during cath	None
293	Line sepsis	Discharge with elevated WBC and fever
294	Femoral artery dissection; aspiration	Over-sedation; failure to monitor patient hemodynamically in ICU
295	None	None
296	Intracranial hemorrhage due to streptokinase	None
297	Occlusion of subtotal circumflex lesion during cath	Periocardiocentesis not attempted
298	Sudden asystolic death	Undetermined
299	Left arm hematoma requiring transfusion, left circumflex artery dissection requiring CABG [X4]	None
301	None	CABG not considered where lesions indicated
302	None	None

QID *	Adverse Event	Quality Problem
303	None	None
304	Small hematoma and mild hematocrit fall	None
305	None	None
309	Anemia requiring transfusion	Inadequate anticoagulation post-tPA
310	None	None
311	Fall in hematocrit after PTCA	None
312	Reocclusion of LAD after PTCA	Discharged with low blood pressure without physical examination
313	None	None
314	None	None
315	Recurrent chest pain after PTCA	None
316	None	None
318	None	None
319	None	None
320	None	None
321	CHF; anemia	Volume overload
322	None	None
323	None	None
325	None	None
326	Anemia	Discharged without addressing anemia
328	None	None

QID *	**Adverse Event**	**Quality Problem**
329	None	None
330	None	None
331	None	None
332	None	None
901	Retroperitoneal hematoma s/p PTCA	None
902	Pericardial tamponade	Perforation of left circumflex artery during PTCA
903	RCA dissection during PTCA	None
904	Renal failure; LIMA dissection	None
905	Acute LAD reocclusion following cessation of heparin	None
906	Reocclusion and small MI post PTCA	None
907	Groin hematoma and drop in hematocrit	Hematocrit not addressed
908	Groin hematoma following PTCA	None
909	None	None
910	Pneumonia	None
911	Cardiac arrest during cath	None
912	None	None

* The QID is the record identification number. Records #201-299 and #901-912 failed screening at JHH; records #301-332 had passed it. Numbering is discontinuous as the leftmost digit identifies the study stratum. Some strata had less than 99 cases. The two rightmost digits of the QID range between 01 and 99; 00 is not used. Review forms for records QID #207, 256, 285, 306, 307, 308, 317, 324, and 327 were unavailable for review, while three additional medical records were unavailable to study reviewers.

Appendix A3. Listing of adverse events and quality problems by record ID, CABG patients (N=213)

QID *	**Adverse Event**	**Quality Problem**
401	Small MI during CABG	None
402	Ruptured diverticulum	None
403	Perioperative CVA	Poor antibiotic choice; poor respiratory management
404	Postoperative bleeding	Poor use of antibiotics
405	Sternal and leg harvest site infection; VF arrest	None
406	Cellulitis	Cellulitis postoperative
407	Arterial puncture during CABG	No daily progress note in chart
408	None	None
409	Neurologic changes	None
410	Stroke	Lack of progress notes 2 days prior to discharge; worsening of creatinine not resolved by discharge; heparin dose error
411	Pneumonia	Cardiac medication without indication
412	None	None
413	None	None
414	Postoperative death	None
415	Sternal wound infection	Poor management of fluid status
416	A-fib	None
417	None	None
418	None	None

QID *	**Adverse Event**	**Quality Problem**
419	Sacral decubitus ulcer	Diabetic, glucose <85 within 2 days of discharge
420	Pneumothorax	None
421	Cardiac arrest	Positive culture of central line
422	Death following MI and CABG	None
423	Perforated diverticulum	None
424	Sepsis	Delay in medical consultation
425	Mediastinitis	Drop in hematocrit not addressed
426	None	None
427	DVT	No postoperative subcutaneous heparin
428	Perioperative MI	Discharged without beta blocker; no post MI risk assessment
429	Postoperative leg infection	Poor management of infection
430	Stroke	None
431	Femoral vein laceration	None
432	Acute renal failure	None
433	Stroke; aspiration; ARDS	No physician note documenting adverse clinical event
434	Pneumothorax	None
435	Hypoxemia requiring intubation; resp. arrest	Pulmonary embolism not considered
436	Death	None
437	Postoperative mediastinal bleeding requiring re-operation	Creatinine significantly higher at discharge

QID *	**Adverse Event**	**Quality Problem**
438	Postoperative mediastinitis	None
439	Intraoperative stroke; wound dehiscence	None
440	Stroke after CABG	None
441	Multiorgan system failure	None
442	Postoperative stroke	None
443	Sternal dehiscence	None
444	Laceration of intercostal artery; multiorgan system failure	None
445	Postoperative respiratory failure	Failure to broaden antibiotic coverage in a timely fashion
446	VT/VF arrest	None
447	Cardiogenic shock	Delay in IABP insertion
448	None	None
449	Stroke	No plan for respiratory treatment after discharge; large bolus of heparin
450	None	None
451	Recurrent ischemia following acute closure of bypass graft	None
452	DVT	None
453	Laceration of old bypass graft	None
454	Postoperative stroke	Poor documentation of home care referral
455	Death	None
456	Postoperative vascular insufficiency	None

QID *	Adverse Event	Quality Problem
457	None	None
458	Postoperative pancreatitis; necrotic IV site	IV site not checked
459	Graft reocclusion	Thyroid supplement without T4 or TSH
460	Postoperative VF arrest	None
461	Pulmonary embolus	Postoperative subcutaneous heparin
462	GN line sepsis	Break in sterile technique
463	Arterial occlusion (needed embolectomy); MI after cath	None
464	Cellulitis	None
465	Renal failure; pneumonia	
466	None	None
467	Bleeding during CABG requiring reexploration	None
468	Postoperative acute renal failure	None
469	Intraoperative MI	None
470	Postoperative VT and death	None
471	Postoperative bleeding needing reexploration	None
472	CHF; ischemia of gut and liver	None
473	Postoperative MI; vocal cord paralysis	None
474	Large stroke; aspiration pneumonia	None
475	Postoperative arrhythmia	None

QID *	Adverse Event	Quality Problem
476	Postoperative bilateral femoral embolus	None
477	Sternal wound infection needing readmission	No physician response to nurse note of erythema in sternum
478	Intraoperative stroke	None
479	A-fib with rapid ventricular response	None
480	Postoperative bleeding requiring reexploration	None
481	Brachial plexus injury	None
482	Intraoperative stroke	None
483	Stroke after CABG	Carotid screening test interpreted by non-MD
484	Intraoperative stroke	Abrupt withdrawal of benzodiazepene in Ativan-dependent patient
485	None	None
486	Fractured tooth due to intubation	None
487	Postoperative VT/VF	
488	None	None
489	None	None
490	Thrombosed vein graft	Failure to assess graft before closing
491	stroke	None
492	Cellulitis in vein harvest site	None
493	None	None
494	None	None

QID *	**Adverse Event**	**Quality Problem**
495	Cardiac arrest	Toxic blood levels of procainamide
496	None	None
497	Postoperative stroke	Decubitus ulcer
498	None	None
499	Postoperative MI	CPK with MB: no action taken
501	None	None
502	None	None
503	None	None
504	A-fib	None
505	None	No medical notes for 2 days before discharge
506	A-fib	Inadequate follow-up
507	Worsening renal failure requiring dialysis	None
508	None	None
509	None	None
510	Lower extremity hematoma	Oozing before discharge not addressed
511	Jaw injury from intubation	Low hematocrit not addressed
512	None	None
513	Postoperative bleeding	No CXR; readmitted with pleural effusion
514	None	None
515	Intra-cardiac foreign body; recurrent VT	Intracardiac foreign body

QID *	**Adverse Event**	**Quality Problem**
516	Hyperkalemia and sudden death	Inadequate monitoring
517	None	None
518	None	None
519	None	None
520	None	None
521	Postoperative bleeding	None
522	Fever presumably 2nd to atelectasis	Preventable with pulmonary toilet
523	Altered mental status postoperatively	None
524	None	None
525	None	Reason for no transfusion not documented
526	Postoperative cholecystitis	None
527	None	None
528	None	None
529	Draining wound at harvest site	Delay in addressing cellulitis
530	Anemia post CABG	Decreasing hematocrit during hospitalization
531	None	Abnormal pO2 not followed up
532	None	None
533	None	None
534	Intra/postoperative MI	Inappropriate insulin therapy
535	None	None

QID *	**Adverse Event**	**Quality Problem**
601	MI after CABG	Possible technical problem in surgery
801	Postoperative tamponade and stroke	None
802	Pulmonary embolism	None
803	Aspiration after extubation	Inappropriate CPT: patient refusal; patient drowsy (Compazine); patient in Trendelenburg position
804	Airway trauma	Multiple unsuccessful attempts to intubate
804	None	Supplemental insulin until discharge but not discharged on insulin
805	Bleeding during/following CABG	None
806	Progressive hypotension and cardiac arrest, required re-do of bypass, IABP	
807	RCA dissection ??? PTCA	None
808	Transfusion-induced hemolytic anemia	None
809	Arterial bleeding from graft	Failure to recognize bleeding promptly
810	Urinary retention	None
811	VT arrest during CABG	None
812	Pleural effusion following CABG/aortic valve replacement	None
813	VF and shock s/p CABG	Increased lidocaine to toxic levels

QID *	Adverse Event	Quality Problem
812	Pleural effusion following CABG/aortic valve replacement	None
813	VF and shock s/p CABG	Increased lidocaine to toxic levels
814	MI/arrest following CABG	Inadequate sedation; inadequate exploration of CP
815	Sternal infection post-CABG	None
816	Hematoma and fall in hematocrit post-CABG	None
817	Arterial thrombus on cath site; aspiration	Delay in detection of NG tube misplacement
818	Allergy to streptokinase	None
819	Post-angiogram hypotension: CCU transfer	None
820	Altered mental status following CABG	None
821	None	None
822	None	None
823	Intraoperative stroke	Unclear
824	Sternal wound infection during CABG	None
825	None	None
826	Sepsis and wound dehiscence	None
827	VF arrest and complete heart block post-CABG	None
828	Intraoperative MI following VF	None
829	Arrest after bypass	None

QID *	Adverse Event	Quality Problem
830	Postoperative respiratory distress; transient renal insufficiency following bypass	Lack of follow-up
831	Postoperative stroke; allergy to penicillin	Allergy to penicillin reported, given ampicillin
832	Wound dehiscence after CABG; postoperative arrhythmias	None
833	Respiratory arrest; renal failure; pericarditis	None
834	Postoperative urinary retention; postoperative bleeding	Unsecured branch of LIMA requiring emergent re-operation
835	Postoperative bleeding causing tamponade	Problem in surgical technique
836	Postoperative stroke and respiratory distress	None
837	Stroke during CABG	None
838	Postoperative heart block	Incorrect (not ordered) medication
839	Bilateral pneumothorax after removal of mediastinal tube	Removal (premature?) of chest tube requiring reinsertion
840	Sternal wound dehiscence; postoperative respiratory failure	Lack of documentation of several ICU days
841	Complete heart block during cardiac cath	None
842	None	None
843	Postoperative pneumonia/sepsis	None
844	Postoperative graft thrombosis and MI	None
845	Postoperative stroke; aortic tear	None

QID *	**Adverse Event**	**Quality Problem**
846	Post/intraoperative stroke	None
847	Shearing of catheter tip	Shearing of catheter tip
848	Aortic rupture	None
849	Postoperative foot drop	None
850	Vein harvest site dehiscence	Surgical drain sutured to wound closure
851	None	None
852	Foot drop; altered mental status	No physician notes last 3 days
853	Hand weakness and paresthesia	None
854	None	No physician notes prior to discharge
855	Injury during intubation	None
856	AV fistula from cath	None
857	Pulmonary edema post-transfusion	None
858	None	None
859	Mild hand weakness from TIM	Incorrect heparinization
860	Atrial and ventricular arrhythmia; arrest	Failure to monitor patient
861	VT	Failure to intubate electively
862	None	None
863	Renal failure	None
864	Sternal wound infection; pulmonary failure	Inadequate follow-up
865	Disseminated intravascular coagulation	None

QID *	**Adverse Event**	**Quality Problem**
868	Scalp abrasion/cellulitis	None
869	Hematuria after Foley catheter insertion	None
870	None	None
871	Sepsis	None
872	None	None
873	None	None
874	Rotator cuff tear following CT	Delay in diagnosis of cuff tear
875	Stroke during CABG	Carotid occlusion not explored adequately
876	Intraoperative stroke	None
877	None	Elevation of creatinine/BUN not discussed in notes

* The QID is the record identification number. Records #401-499, #601, and 80l-877 had failed screening indicators at JHH; records #501-537 had passed it. Numbering is discontinuous as the leftmost digit identifies the study stratum. Some strata had less than 99 cases. The two rightmost digits of the QID range between 01 and 99; 00 is not used.

Abbreviations used in Appendix tables

A-fib	atrial fibrillation
ABG	arterial blood gas measurement
AMI	acute myocardial infarction
ARDS	acute respiratory distress syndrome
AV	aspirin
cath	arteriovenous
CCU	cardiac catheterization
CHF	coronary care unit
CPK	creatine phosphokinase enzyme level
CPT	chest physiotherapy
CT	computed tomography
CXR	chest x-ray
DVT	deep venous thrombosis
ECG	electrocardiogram
EMD	electromechanical dissociation
GI	gastrointestinal
IABP	intraaortic balloon pump
ICU	intensive care unit
IV	intravenous
lab	laboratory
LCA	left coronary artery
LIMA	left internal mammary artery
MB	myocardial band (of CPK enzyme)
MI	myocardial infarction
NG	nasogastric
pO2	partial pressure of oxygen
PTT	partial thromboplastin time
RCA	right coronary artery
s/p	status post
T4	thyroxin
tPA	tissue plasminogen activator (thrombolytic)
TSH	thyroid stimulating hormone
UTI	urinary tract infection
VF	ventricular fibrillation
VT	ventricular tachycardia
WBC	white blood cell count

CMS LIBRARY
3 8095 00012700 7